Gastrointestinal Disease Detection using Transformer-Based Image Segmentation

Nevasini Sasikumar [1] Krishna Sri Ipsit Mantri [2]

[1]PES University [2]Indian Institute of Technology Bombay

ABSTRACT

These days gastrointestinal diseases such as ulcer, polyp, bleeding are wide spreading and common. Manual diagnosis is time consuming and in-accurate, aiding to this fact we propose a simple lightweight deep learning transformer based approach that can effectively perform image segmentation of the gastrointestinal tract. By effective we aim at reducing the FLOPs by >85 percent. Our method captures both feature representations and specific patch embeddings obtained from the transformer model. Using Knowledge distillation we can pass on information from the parent to the student model catering to learning from multiple points within the transformer architecture. Every pixel of the teacher features is transferred to all the spatial and semantic locations of the student.

Method

This work is mainly divided into four fundamental modules and each module performs its unique operation none of then being interdependent.

- Patch Embedding Distillation (PED)
- Universal and Local Relation Mixer (UL-Mixer)
- Auxiliary Latent Representation Assistant (ALRA)
- Cross Channel Blend (CCB)

Detailed Architecture

The entire framework has been sub-divided into namely 4 modules

- **Patch Embedding Distillation**: The patch-group distillation that allows the student to learn the local spatial feature from patches and retain the correlation among them. Given the original student feature F^S and teacher feature F^T they are partitioned into $n \times m$ patches of size $h \times w$, where $h = H/n$, $w = W/m$. They are further arranged as g groups sequentially where each group contains $p = n \cdot m/g$ patches. Specifically, the patches in a group will be concatenated channel-wisely, forming a new tensor of size $h \times w \times c \cdot g$ that would be used for distillation lately.

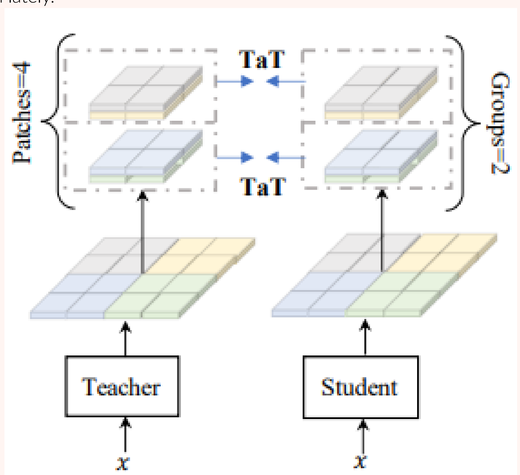


Figure 1. Patch Embedding Distillation

- **UL-Mixer**: We take into consideration both the local and global feature representation and entail the corresponding interrelation among them. The global context of an embedding is extracted by multi-head attention layer, whereas the local features are extracted by convolution operations.

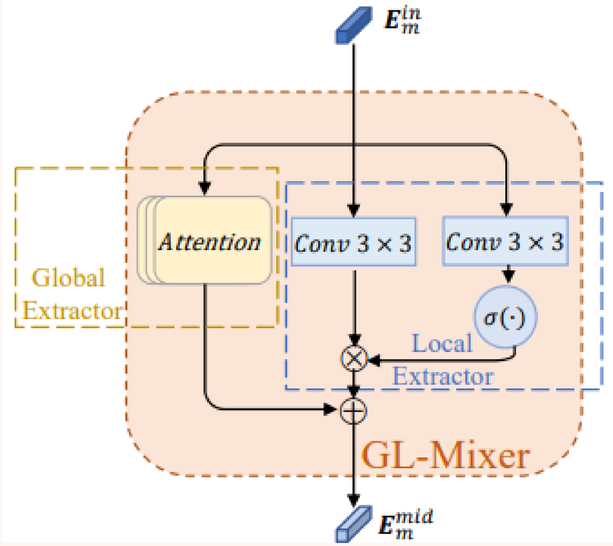


Figure 2. UL Mixer

- **Auxiliary Latent Representation Assistant**: Latent Representation-based Assistant knowledge distillation using intermediate-sized assistant models was introduced to alleviate the poor learning of a student network when the size gap between a student and a teacher is large. It achieved an effective performance improvement in the case of a large gap in teacher and student sizes.

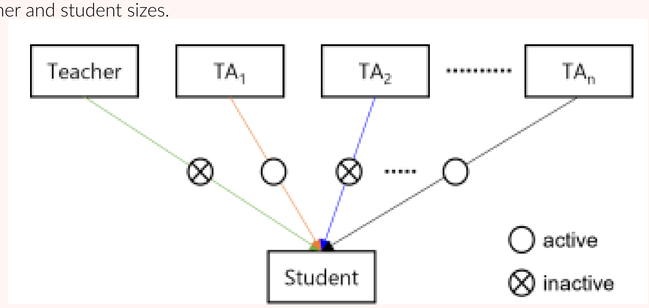


Figure 3. Auxiliary Latent Representation Assistant

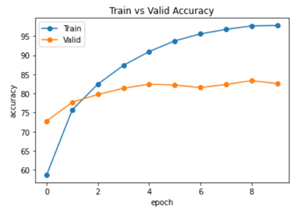
Dataset

We use the open-access dataset of gastrointestinal polyp images and their corresponding segmentation masks. This dataset contains 1000 polyp images with resolutions varying from 332×487 to 1920×1072 pixels, and their corresponding 1-bit masks indicating the presence of polyps.

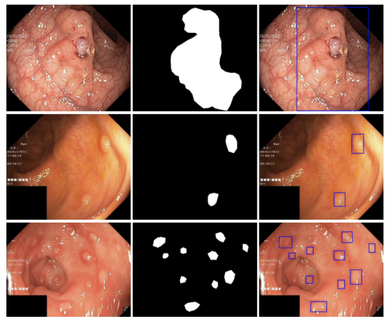
(a) Image (b) Mask

Results

We trained the model using 900 images and tested it using the remaining 100 images.



(a) Accuracy Plot



References

[1] Sihao Lin, Hongwei Xie, Bing Wang, Kaicheng Yu, Xiaojun Chang, Xiaodan Liang, and Gang Wang. Knowledge distillation via the target-aware transformer, 2022.

[2] Ruiping Liu, Kailun Yang, Alina Roitberg, Jiaming Zhang, Kunyu Peng, Huayao Liu, and Rainer Stiefelhagen. Transkd: Transformer knowledge distillation for efficient semantic segmentation, 2022.